REMARKS

Status of the Claims

Claims 26, 30 and 31 are pending in the application. Claims 26, 30 and 31 are rejected. Claim 31 is canceled and claims 26 and 30 are amended herein. No new matter is added to the amended claims.

Claim amendments

Claim 31 is canceled and its limitation incorporated in Claim 26 to overcome the rejection under 35 U.S.C. §112, first paragraph. Amended claim 26 is drawn to a method of producing activated T cells directed towards stratum corneum chymotrytic enzyme (SCCE) in an individual with ovarian cancer, prostate cancer or breast cancer. Such a method comprises the steps of exposing dendritic cells to a human stratum corneum chymotrytic enzyme peptide selected from the group consisting of SEQ ID NOs: 31, 32, 33, 34, 35, 36, 80, 86, 99 or the human stratum corneum chymotrytic enzyme protein encoded by the DNA of SEQ ID NO: 30, thereby producing activated dendritic cells. The activated dendritic cells are then exposed to T cells, where the activated dendritic cells would present the stratum corneum chymotrytic enzyme peptide to the T cells, thereby producing activated T cells directed toward the stratum corneum chymotrytic enzyme in the individual with ovarian cancer, prostate cancer or breast cancer.

Claim 30 is amended to properly depend from amended claim 26. Amended claim 30 recites the step where the dendritic cells are isolated from the

individual prior to the exposure to the stratum corneum chymotrytic enzyme polypeptide, where the activated dendritic cells are reintroduced into the individual subsequent to the exposure.

The 35 U.S.C. §112, First Paragraph Rejection

Claims 26 and 30-31 remain rejected under 35 U.S.C. §112, first paragraph for lack of enablement. Applicant respectfully traverses this rejection.

The Examiner states that the Declarations filed on 2/19/2003, 1/23/2006 and 9/1/2006 are insufficient to overcome the rejection of claims 26, 30 and 31 based upon the insufficiency of the disclosure under 35 U.S.C. §112, first paragraph because the showing is not commensurate in scope with the claims. Additionally, the Examiner states that although the Declarations submitted provide evidence that stratum corneum chymotrytic enzyme peptides with SEQ ID NOs: 32 and 33 can induce CTL response *in vitro*, the claims still encompass stratum corneum chymotrytic enzyme peptides other than SEQ ID NOs: 32 and 33 that do not contain amino acids 5-13 or 123-131 of SCCE as well as the treatment of just any disorder, particularly ovarian, prostate, breast and colon cancer as recited in claim 31.

Furthermore, the Examiner states that although the instant claims are drawn to stratum corneum chymotrytic enzyme dendritic cell immunotherapy in cancer patients, the Declarations are limited to *in vitro* evidence and nucleic acid expression rather than polypeptide expression. The Examiner then cites teachings of **Fu** *et al*, **Powell** *et al*, **Vallejo** *et al* and **Lewin B** to state that

analysis of levels of polynucleotide transcripts cannot be relied upon to anticipate levels of protein expression and one of skill in the art cannot anticipate that the level of a specific mRNA expressed by a cell will be paralleled at the protein level due to complex homeostatic factors controlling translation and post-translational modification.

With regards to the references submitted by the Applicant in the last response, the Examiner states that with exception to the art of **Noriaki** *et al*, **Higuchi** *et al* and **Van Gluck** *et al*, the other references are limited to *in vitro* stimulation of dendritic cells and provide a starting point for further experimentation into their use for patient specific immunotherapy. While pointing out to the fact that **Higuchi** *et al* and **Van Gluck** *et al* were published after the filing date of the present application, the Examiner states that the claims still encompass stratum corneum chymotrytic enzyme peptides other than SEQ ID Nos: 32 and 33 that do not contain amino acids 5-13 or 123-131 of stratum corneum chymotrytic enzyme and there is no evidence that the stratum corneum chymotrytic enzyme signal peptide is part of the mature stratum corneum chymotrytic enzyme protein expressed in cancer patients such that the stratum corneum chymotrytic enzyme activated T cells targeting this sequence effectively inhibit tumor growth in a patient.

The Examiner further reiterates that in view of unpredictability in the art to which the invention pertains as evidenced by **Cranmer** *et al*, **Soruri** *et al*, **Wang** *et al* and **Geysen** *et al*, the lack of established clinical protocols for effective dendritic cell therapies, undue experimentation would be required to

practice the claimed methods with a reasonable expectation of success, absent a specific and detailed description in the instant specification of how to effectively practice the claimed methods and absent working examples providing evidence which is reasonably predictive that the claimed methods of dendritic cell immunotherapy to treat ovarian and prostate cancer patients in particular, commensurate in scope with the claimed invention. For these reasons, the Examiner has maintained rejection of claims 26 and 30-31 for lack of enablement.

Claim 26 is amended as discussed supra and is drawn to a method of producing activated T cells directed towards stratum corneum chymotrytic enzyme in an individual with ovarian cancer, prostate cancer or breast cancer. The inclusion of breast cancer is supported by the teaching in the art (Fig. 6, Harvey et al., 2000, *J Biol Chem*, 275(48): 37397-37406) and in the Declaration included with the response filed August 4, 2006, both of which disclose expression of stratum corneum chymotrytic enzyme in breast cancer.

With regard to the Examiner's contention that there was no evidence of induction of CTL response provided for peptides other than the ones with SEQ ID NOs: 32 and 33, Applicant would like to respectfullly point out the patentability rules, which state that "the specification need not contain an example if the invention is otherwise disclosed in such a manner that one skilled in the art will be able to practice it without undue experimentation. Lack of working example, however, is a factor to be considered, especially in a case involving an unpredictable and undeveloped art. But because an enabling disclosure is required, Applicant need not describe **all actual embodiments**" (M.P.E.P.

2164.02). Since the Declarations enable the claimed method for peptides with SEQ ID Nos: 32 and 33, Applicant submits that a person having ordinary skill in this art will be able to practice the claimed method using other peptides without undue experimentation.

In response to the Examiner's statement that analysis of levels of polynucleotide transcripts should not be relied upon to anticipate levels of protein expression, Applicant states that all the techniques mentioned herein are well known and acceptable in the art and are routinely used by investigators to examine the expression of genes in several disorders. Given this, one skilled in the art can easily examine expression of proteins in these cancers using techniques that are routine in the art, for instance, western blots, immunohistochemical analysis, ELISA etc.

Further, in response to Examiner's statement that the evidence provided by the Applicant with regards to dendritic cell immunotherapy were mostly limited to *in vitro* experiments, Applicant would like to respectfully point out that *in vitro* studies are accepted by those with ordinary skill in the art of gene therapy and anticancer therapy as being predictive of success *in vivo*. Furthermore, the Examiner has not cited any case law stating that the 35 U.S.C. §112 requires *in vivo* or human clinical data for Applicant's method to satisfy the enablement requirement Nevertheless, as stated by the Examiner, **Noriaki** *et al* demonstrated efficacy of using dendritic cell immunotherapy in patients. Thus, dendritic cell immunotherapy is not as unpredictable as stated by the Examiner.

Additionally, with regard to the issue of whether stratum corneum chymotrytic enzyme signal peptide is expressed in cancer patients, Applicant states that it is not necessary to show that the signal peptide is expressed in cancer patients. The most relevant demonstration is of the immunogenic potential of the signal peptide. The Declaration provided with the response filed January 20, 2006 demonstrated that the peptide with SEQ ID NO: 33 that is derived from a signal peptide sequence was capable of inducing a CTL response to lyse CaOV3 ovarian tumor cells that were not pulsed with the peptide. This data provides a strong evidence that the peptide with SEQ ID NO: 33 may be naturally expressed on the ovarian tumor cells. Since the enablement requirement does not require that all the embodiments be described, Applicant contends that the *in vitro* data provided by the Applicant and the references provided by the Applicant demonstrating the efficacy of dendritic cell immunotherapy *in vitro* and *in vivo* are sufficient to enable the instantly claimed methods.

Applicant reiterates as stated in In re Woody, 331 F.2d 636, 639, 141 USPQ 518, 520 (CCPA 1964): "It appears that no one on earth is certain as of the present whether the process claimed will operate in the manner claimed. Yet absolute certainty is not required by the law. **The mere fact that something has not previously been done clearly is not, in itself, a sufficient basis for rejecting all applications purporting to disclose how to do it.**" A novelty in the claimed invention relates in part to the identification of immunogenic stratum corneum chymotrytic enzyme peptides and not to the art of loading the dendritic cells with these peptides. Nevertheless, the induction of cytotoxic T cell response

by some of the claimed stratum corneum chymotrytic enzyme peptides is sufficient to demonstrate the success in loading of dendritic cells with these peptides, culturing the loaded dendritic cells and production of immune activated T cells. Therefore, a fair reading of the instant specification along with the information available in the art should enable one of skill in the art to make and use the claimed invention without undue experimentation. Thus, Applicant submits that the scope of the claimed invention is commensurate with the enablement provided. Based on the above-mentioned amendments and remarks, Applicant respectfully requests the withdrawal of rejection of claims 26 and 30-31 under 35 U.S.C. §112, first paragraph.

Claim 31 stands rejected under 35 U.S.C. §112, first paragraph as failing to comply with the written description requirement as introducing new matter. Applicant respectfully traverses this rejection.

The Examiner states that the response filed by the Applicant on September 1, 2006 introduced new matter into the claims. As presently amended claim 31 is drawn to the claimed immunotherapeutic method of wherein the individual has breast or colon cancer. The Examiner further states that although the PTO has the initial burden of presenting evidence or reasons why persons skilled in the art would not recognize description of the invention defined by the claims in the disclosure, when filing an amendment the applicant should show support in the original disclosure for new or amended claims. Hence, the Examiner states that a sufficient written support in the specification or claims as

filed should be provided by the Applicant for the limitations recited in the presently amended claim 31 or these limitations must be removed from the claims in response to this Office Action.

Claim 31 is canceled herein. Additionally, amended claim 26 is drawn to a method of producing activated T cells directed towards stratum corneum chymotrytic enzyme in an individual with ovarian, prostate or breast cancer. The limitation of "colon cancer" is not included in amended claim 26. Applicant submits that expression of stratum corneum chymotrytic enzyme in breast cancer was known in the art (Fig. 6, Harvey et al., 2000, *J Biol Chem*, 275(48): 37397-37406) at the time the invention was filed and that one skilled in the art would easily recognize that the instantly claimed method would also be useful in inducing an immune response in breast cancer. Furthermore, the declaration provided by the Applicant along with the response filed August 4, 2006 had data that was obtained prior to the filing of the instant invention (September 9, 1997) and further confirmed the expression of stratum corneum chymotrytic enzyme in breast cancer. Hence, Applicant contends that the inclusion of breast cancer in the amended claim 26 did not introduce any new concept. Accordingly, based on this amendment and remarks, Applicant respectfully request the withdrawal of rejection of claim 31 under 35 U.S.C. §112, first paragraph.

This is intended to be a complete response to the Office Action mailed November 17, 2006. Applicant submit that the pending claims are in

condition for allowance. If any issues remain outstanding, please telephone the undersigned attorney of record for immediate resolution.

Respectfully submitted,

Date: 5/17/07

Benjamin Aaron Adler, Ph.D., J.D.
Registration No. 35,423
Counsel for Applicant

ADLER & ASSOCIATES
8011 Candle Lane
Houston, Texas 77071
713-270-5391
badler1@houston.rr.com

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 275, No. 48, Issue of December 1, pp. 37397–37406, 2000
Printed in U.S.A.

Tissue-specific Expression Patterns and Fine Mapping of the Human Kallikrein (*KLK*) Locus on Proximal 19q13.4*

Received for publication, May 25, 2000, and in revised form, August 31, 2000
Published, JBC Papers in Press, August 31, 2000, DOI 10.1074/jbc.M004525200

Tracey J. Harvey‡, John D. Hooper‡, Stephen A. Myers‡, Sally-Anne Stephenson‡§, Linda K. Ashworth¶, and Judith A. Clements‡‖

From the ‡Centre for Molecular Biotechnology, School of Life Sciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia, and the ¶Lawrence Livermore National Laboratory, Livermore, California 94551

The tissue or glandular kallikreins (*KLK*) are members of a highly conserved multigene family encoding serine proteases that are central to many biological processes. The rodent *KLK* families are large, highly conserved and clustered at one locus. The human *KLK* gene family is clustered on chromosome 19q13.3–13.4, and until recently consisted of just three members. However, recent studies have identified up to 11 new members of the *KLK* family that are less conserved than their rodent counterparts. Using a Southern blot and sequence analysis of 10 BACs and cosmids spanning approximately 400 kilobases (kb) either side of the original *KLK* 60-kb locus, we demonstrated that these genes also lie adjacent to this. We have also clarified the position of several microsatellite markers in relation to the extended *KLK* locus. Moreover, from Southern blot analysis of the cosmids and BACs with a degenerate oligonucleotide probe to the histidine-encoding region of serine proteases, we have shown that there are no other serine protease genes approximately 400 kb centromeric and 220 kb telomeric of the extended locus. We performed an extensive analysis of the expression patterns of these genes by poly(A)$^+$ RNA dot blot and reverse transcriptase-polymerase chain reaction analysis, and demonstrated a diverse pattern of expression. Of interest are clusters of genes with high prostate (*KLK2–4*) and pancreatic (*KLK6–13*) expression suggesting evolutionary conservation of elements conferring tissue specificity. From these findings, it is likely that the human *KLK* gene family consists of just 14 clustered genes within 300 kb and thus is of a comparable size to the rodent families (13–24 genes within 310 and 480 kb, respectively). In contrast to the rodent families, the newest members of the human *KLK* family are much less conserved in sequence (23–44% at the protein level) and appear to consist of at least four subfamilies. In addition, like the rat, these genes are expressed at varying levels in a diverse range of tissues although they exhibit quite distinct patterns of expression.

Serine proteases are a group of protein-cleaving enzymes that contain a serine residue in their active site and play important roles in diverse physiological processes. The glandular or tissue kallikreins are a subgroup of serine proteases highly conserved across several species, that are involved in the post-translational processing of polypeptides to their bioactive or inactive forms, a function that underlies and is central to most biological events (1). Tissue kallikrein (encoded by *KLK1*),[1] from which this family derives its generic name, processes kininogen to release (lys)-bradykinin, a multifunctional peptide involved in the regulation of local blood flow, sodium balance, inflammation, and cell proliferation in many tissues (2). Other members of the *KLK* family include growth factor processing enzymes such as mouse epidermal growth factor-binding protein, γ-nerve growth factor, and enzymes that generate vasoactive peptides such as the mouse pro-renin converting enzymes, γ-renin, and rat tonin (3–5). In man, the two other most well characterized enzymes, prostate-specific antigen, PSA (encoded by *KLK3*; Ref. 6), and hK2 (encoded by *KLK2*; Ref. 7), act on entirely different substrates. Nevertheless, each is involved in the selective cleavage of specific precursors to their bioactive or inactive forms (8). These findings highlight the involvement of the KLK family in a number of different and complex enzymic pathways that underlie several important biological functions.

The wide range of potential functions of these enzymes is emphasized by their diverse, yet specific, tissue expression patterns. Whereas *KLK2* and *KLK3* are primarily expressed at high levels in the secretory epithelium of the prostatic ducts and have been implicated in normal prostatic function as well as in cancer progression (8), tissue kallikrein, through its multifunctional roles, is involved in the (patho)physiology of the kidney, brain, and respiratory, gastrointestinal, and reproductive tracts (2). This diverse functionality is underscored by the ubiquitous expression pattern observed for *KLK1* in the rat (9). Although a function has not yet been ascribed to the protein products of most other rat *KLK* genes, an indication of the tissue in which they will be functional is gained from their tissue expression patterns (9). In man, the newly identified KLK serine proteases, Protease M/neurosin/zyme (10–12) and normal epithelial cell-specific 1 (Nes1)[2] (13), are highly ex-

* This work was supported by the National Health and Medical Research Council of Australia, Australian Research Council, Queensland Cancer Fund, Queensland University of Technology, and U.S. Department of Energy contract W-7405-Eng-48 to the Lawrence Livermore National Laboratory. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Current address: Haematology-Oncology Dept., The Queen Elizabeth Hospital, Woodville, SA, 5011 Australia.

‖ To whom correspondence should be addressed: Centre for Molecular Biotechnology, School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD, Australia 4001. Tel.: 61-7-38641899; Fax: 61-7-38641534; E-mail: j.clements@qut.edu.au.

[1] Approved nomenclature by the Human Genome Nomenclature Committee for the human kallikrein family: ***KLK1***, ***KLK2***, ***KLK3***, ***KLK4***:Prostase KLKL1; ***KLK5***:KLKL2; ***KLK6***:Protease M; ***KLK7***: SCCE; ***KLK8***:Neuropsin; ***KLK9***:KLKL3; ***KLK10***:Nes1; ***KLK11***:TLSP; ***KLK12***:KLKL5; ***KLK13***:KLKL4; ***KLK14***:KLKL6. Since this nomenclature is very recent, old and new gene designations will be used concurrently to avoid confusion. Details of the new nomenclature can be found on the Human Genome Nomenclature Committee Website and in Diamandis *et al.* (50).

[2] The abbreviations used are: Nes1, normal epithelial cell-specific 1;

pressed in brain and breast tissue, and have been suggested to be involved in Alzheimer's disease and breast cancer, respectively (12–14). Several other recently identified *KLK*-like serine protease genes have also been shown, at the RT-PCR level, to be expressed in a tissue-specific manner, which presumably indicates the tissues in which they will be biologically active (15–20).

The *KLK1*, *KLK2*, and *KLK3* genes are clustered within a region that spans 60 kb on chromosome 19q13.3-13.4 (21, 22). This cluster maps to a position ~55.5 Mb from the p-telomere on the metric physical map of the chromosome and at the proximal end of 19q13.4. The larger size of the rat and mouse *KLK* gene families (13–24 genes) (23–25), which are also clustered in one locus (23, 25, 26), and the recent identification of a new *KLK* gene, *KLK4* (prostase/KLKL1) (15, 21, 27) 25 kb downstream of the *KLK2* gene (15, 21), suggested that the human *KLK* gene family would be larger than previously thought. Indeed, Southern blot analysis of a human genomic library, using a monkey *KLK* cDNA with 92–95% homology to human *KLK* cDNAs (28), indicated that there may be as many as 19 human *KLK* genes (29). Furthermore, a number of *KLK*-related serine protease gene sequences have been detected that lie telomeric to the original *KLK* locus, thus expanding the human *KLK* gene family (30). However, the physical mapping of this locus has not yet been described and the definitive size of the family is not clear. In addition, the location of the various microsatellite markers in the *KLK* locus is not clear.

In this study, we have physically mapped the serine protease genes clustered at the *KLK* locus and defined the position of key microsatellite markers. We have also confirmed the order and orientation of the 10 putative new members of the human *KLK* family which span approximately 300 kb across the *KLK* locus on chromosome 19q13.4. We have further determined that no other serine protease genes are present within the regions surrounding the *KLK* locus. In addition, extensive gene expression studies have identified the most abundant sites of expression for these genes and a diverse tissue specificity of expression similar to the rodent *KLK* family. Also, phylogenetic analysis of the 14 serine proteases encoded by the *KLK* genes of this locus demonstrated the extent of divergence within the expanded family and suggested that there are at least four human *KLK* subfamilies.

EXPERIMENTAL PROCEDURES

Cosmid and BAC DNA Clones—Four cosmids (R32740, R31381, F22702, and R28781) and six BAC clones (BC772576, BC778306, BC33747, BC85745, BC349237, and BC892989), that spanned the classical *KLK* locus (*KLK1–3*) and approximately 400 kb either side of this region, were used in the following studies. Five hundred-ml cultures of Luria-Bertani medium supplemented with 20 μg/ml kanamycin or 15 μg/ml chloramphenicol, were inoculated with the BAC and cosmid clones, respectively, and incubated overnight at 37 °C. DNA was then extracted using the alkaline lysis method, followed by RNase A digestion and ethanol precipitation.

Physical Mapping of the KLK Locus on Chromosome 19q—Twenty μg of DNA from the set of 10 BAC and cosmid clones were digested overnight at 37 °C with the restriction enzyme *Eco*RI, then separated electrophoretically on a 1% agarose gel. The gel was treated with 1.5 M NaCl, 0.5 M NaOH for 30 min to denature the DNA, rinsed briefly in distilled water, followed by a 30-min incubation in neutralization buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4). The DNA was capillary blotted onto Hybond N™ nylon membrane (Amersham Pharmacia Biotech) overnight in 20 × SSC then cross-linked to the membrane by ultraviolet irradiation. Membranes were hybridized with a ^{32}P-random labeled cDNA probe (Ambion), or a ^{32}P-end labeled gene-specific oligonucleotide probe (Table I), or a ^{32}P-end labeled degenerate oligonucleotide probe to the highly conserved regions spanning either the catalytic histidine codon (GGTSCTSACAGCYGCCCAYTG; 16-fold degeneracy; S = G or C, Y = C or T) or the the catalytic serine codon (CAGIGGICCIC-CIGARTCAAC; 512-fold degeneracy; I = inosine, *r* = A or G) characteristic of serine proteases. Hybridization was performed overnight in ExpressHyb™ solution (CLONTECH) at 42 or 55 °C for oligonucleotide and cDNA probes, respectively. Blots were washed in 1 × SSC, 0.5% SDS up to 55 °C, then exposed to x-ray film (AGFA Curix Ortho) at −80 °C.

TABLE I
Southern blot probes

Southern blot analysis of chromosome 19q cosmid/BAC *Eco*RI-digested DNA was performed using ^{32}P-end labeled oligonucleotide probes for *KLK5–8* and *KLK10–12*, and ^{32}P random-labeled cDNA probes for *KLK9* and *KLK13–14*. Southern blot analysis of RT-PCR products was performed using a digoxigenin-labeled oligonucleotide probe. The percent difference in sequence between an individual probe and each of the 14 genes for all probes used averaged 70%.

Gene	DNA/oligonucleotide probes
KLK1/*KLK1*	5′-CTGGACGTGGGCTTTTTCGCAC-3′
KLK2/*KLK2*	5′-TCTAAGGCTTTGATGCTTCAG-3′
KLK3/*KLK3*	5′-GGTATTTCAGGTCAGCCACAG-3′
KLK4/*KLK4*	5′-CTACCGTGCTGCAGTGCGTG-3′
KLK5/*KLKL2*	5′-AGTGCACTTGGGGGCTCTTGGTT-3′
KLK6/protease M	5′-TGTCCTGGTCATGGCTGGC-3′ (RT-PCR) 5′-AATCACCATCTGCTGTCTTGC-3′
KLK7/SCCE	5′-GCCCAGGTGCACGGTGTACTC-3′
KLK8/Neuropsin	5′-CAATGATCTGCCAGGCTGA-3′ (RT-PCR) 5′-GAGCCCCAGGATGTGATGCCC-3′
KLK9/KLKL3	RT-PCR product (678 bp) 5′-CCCCCAGCCTGAGATGAGACACT-3′
KLK10/Nes1	5′-GCCGTGGTGCCCCAGCCAGCA-3′ (RT-PCR) 5′-GGTAAACACCCCACGAGAGGA-3′
KLK11/TLSP	5′-GGTGATGGAGACTGGCGA-3′
KLK12/KLKL5	5′-AGGTCGTGCTCGTGGCTCGTC-3′
KLK13/KLKL4	EST clone AA002101 *Eco*RI/*Hin*dIII digest (1500 bp) 5′-AGGGGTTAGGCGGTTGTTGTGG-3′
KLK14/KLKL6	EST clone AI935298 *Eco*RI/*Hin*dIII digest (750 bp) 5′-TGTTCACGCATTGCAGAGAGG-3′

Computer Analysis of Chromosome 19 Sequences—Approximately 320 kb of draft sequence of chromosome 19q13.3-13.4 (available at the Lawrence Livermore National Laboratory and the DOE Joint Genome Institute website), beginning ~50 kb upstream of where two of the original *KLK* genes (*KLK2–3*) were localized and searched with the conserved peptide sequences spanning the catalytic histidine, aspartate, and serine residues of *KLK* serine proteases (WVLTAAHC, HDLMLLKL, and GDSGGPL; catalytic residue underlined). This was performed using the data base creating tool and the tBLASTN algorithm available at the Australian National Genome Information Service. In addition, the cDNA sequence of each of the serine proteases, *KLK1* (accession number X13561), *KLK2* (accession number S39329), *KLK3* (accession number M26663), *KLK4* (accession number AF113141), *KLK5*[1] (KLKL2, accession number AF135028), *KLK6* (Protease M, accession number U62801), *KLK7* (SCCE, accession number L33404), *KLK8* (neuropsin, accession number AB009849), *KLK9* (KLKL3, accession number AF135026), *KLK10* (Nes1, accession number AF024605), *KLK11* (TLSP, accession number AB012917), *KLK12* (KLKL5, accession number AF135025), *KLK13* (KLKL4,accession number AF135024), and *KLK14* (KLKL6, accession number AF161221), was aligned against this genomic sequence using the BLASTN algorithm available at the National Center for Biotechnology Information website. BLASTN searches were also performed against the draft chromosome 19q13.3-13.4 sequence using microsatellite marker sequences between 272.77 and 276.46cR$_{3000}$ at interval D19S425-D19S418. Marker sequences used were: WI-9055 and M21896, U37672, stSG39975, sts-S39329, WI-20869, stSG30247, stSG32045, stSG1457, stSG3375, stSG372, sts-H71236, SGC35527, U37672, stSG30429, sts-AA028917, and L12214. The computer software DNAStar (Lasergene) was used to construct an *Eco*RI restriction enzyme map of this sequence. These data were then used in conjunction with the data obtained by physical mapping and the *Eco*RI map available at the Lawrence Livermore National Laboratory website to determine the exact position of each gene and to reconstruct a definitive *Eco*RI map of this region.

Sequence and Phylogenetic Analysis—Multiple alignment of deduced protein sequences was performed using the Eclustal tool available at

RT-PCR, reverse transcriptase-polymerase chain reaction; kb, kilobase(s); PSA, prostate-specific antigen.

TABLE II
PCR primers

Primer pairs used for the detection of the 14 *KLK* genes by RT-PCR and for the generation of DNA fragments for *KLK6–11*, used in poly(A)$^+$ RNA dot blot analyses. The percent difference in sequence between an individual primer and each of the 14 genes for all primers used averaged 74%.

Gene	PCR primers	Product size
KLK1/*KLK1*	Forward 5′-TGGAGAACCACACCCGCCAAG-3′ Reverse 5′-ACGGCGACAGAAGGCTTATTG-3′	468 bp
KLK2/*KLK2*	Forward 5′-GCCTAAAGAAGAATAGCCAGGT-3′ Reverse 5′-CTCAGACTAAGCTCTAGCACAC-3′	375 bp
KLK3/*KLK3*	Forward 5′-GCATCAGGAACAAAAGCGTGA-3′ Reverse 5′-CCTGAGGAATCGATTCTTCAG-3′	138 bp
KLK4/*KLK4*	Forward 5′-GCGGCACTGGTCATGGAAAAGG-3′ Reverse 5′-CAAGGCCCTGCAAGTACCCG-3′	526 bp
KLK5/KLKL2	Forward 5′-GAGCTGGGGCCGGGGAAGAC-3′ Reverse 5′-TGGGCCGGGCACAAGGGTAA-3′	657 bp
KLK6/Protease M	Forward 5′-GAGCGGCCATGAAGAAGC-3′ Reverse 5′-AATCACCATCTGCTGTCTTGC-3′	459 bp
KLK7/SCCE	Forward 5′-GCCCAGGGTGACAAGATTATT-3′ Reverse 5′-GTACCTCTGCACACCAACGG-3′	569 bp
KLK8/Neuropsin	Forward 5′-TACTCTGTGGCGGTGTCCTTG-3′ Reverse 5′-GAGCCCCAGGATGTGATGCCC-3′	523 bp
KLK9/KLKL3	Forward 5′-GGCCGGCCTCTTCCACCTTAC-3′ Reverse 5′-GCGCGGGCTCAGTTCTCCAT-3′	678 bp
KLK10/Nes1	Forward 5′-GCGGAAACAAGCCACTGTGGG-3′ Reverse 5′-GGTAAACACCCCACGAGAGGA-3′	486 bp
KLK11/TLSP	Forward 5′-CCGCTACATAGTTCACCTGG-3′ Reverse 5′-AGGTGTGAGGCAGGCGTAACT-3′	284 bp
KLK12/KLKL5	Forward 5′-TGGCAGACAAAGAGACAAGGT-3′ Reverse 5′-CTTAGAAGGGCTGGCAGGAG-3′	944/815 bp
KLK13/KLKL4	Forward 5′-CTACACCTGCTTCCCCCACTCTCA-3′ Reverse 5′-GCCGGTCAGGTTGCCCACAT-3′	617 bp
KLK14/KLKL6	Forward 5′-CTGGGCAAGCACAACCTGAG-3′ Reverse 5′-GCATCGTTTCCTCAATCCAGC-3′	517 bp

Australian National Genome Information Service. The percentage identity between each of the 14 kallikreins was calculated using the Homologies tool on the Australian National Genome Information Service website. Phylogenetic studies were also performed via the Australian National Genome Information Service website. A distance matrix between the structures of the proteases was computed from the alignment using the Dayhoff program of the Phylip package (31). The Neighbor-joining method was then used to construct a tree from the distance matrix.

Poly(A)$^+$ RNA Dot Blot Analysis—For *KLK1–4*, ^{32}P-end labeled oligonucleotides were used to probe a CLONTECH Master blot containing poly(A)$^+$ RNA from 50 different human tissues. Hybridizations were performed overnight in ExpressHyb™ solution (CLONTECH) at 42 °C. For *KLK6–11* DNA fragments used in the poly(A)$^+$ RNA dot blot analyses were obtained by RT-PCR using the primers detailed in Table II. RT-PCR was performed as described below for the screening of the cancer cell lines. Amplicons were purified using a PCR purification kit (Qiagen) and ligated into pGEM-T or pGEM-T Easy (Promega). DNA insert orientation was examined by sequencing and restriction enzyme digestion analysis. For *KLK5* and *KLK12–14*, probes were generated from EST clones accession number W07551, accession numbers AI394679, AA002101, and AI935298 (Genome Systems Inc), respectively. Prior to radioactive labeling each construct was linearized before transcription using T7, T3, or SP6 RNA polymerase as appropriate to generate an antisense cRNA. Radiolabeled cRNA probes were generated from these clones by incorporation of [^{32}P]UTP (Geneworks) using a StripEz™ kit (Ambion) and hybridized, in Ultrahyb™ solution (Ambion), to a Multiple Tissue Expression array (CLONTECH) containing poly(A)$^+$ RNA from 76 different human tissues and cell lines at 65 °C overnight. All blots were washed in 0.1 × SSC, 0.1% SDS up to 70 °C, and signals were detected by exposure of the blot to x-ray film (AGFA Curix Ortho) for 1–7 days at −80 °C. Control blots, containing DNA for each of the 14 kallikreins, were included in each hybridization to confirm the specificity of the radiolabeled probes.

Cell Lines and Tissue Samples—The following cell lines were obtained from the American Tissue Culture Collection: T47D (breast cancer cell line); HaCAT (keratinocyte cell line); OVCAR-3 (ovarian cancer cell line); HEC1A, Ishikawa, and KLE (endometrial cancer cell lines); and DU145, PC3, and LNCaP (prostate cancer cell lines). An additional prostate cell line, ALVA41, was obtained from Dr. A. Nakla, St Luke's Hospital, New York. The normal kidney tissue samples and the renal carcinoma cell lines, Caki1 and SN12k1, were a gift from Dr.

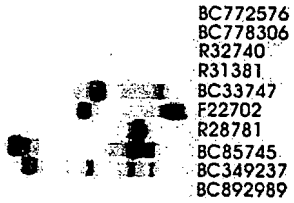

FIG. 1. **Identification of *KLK*-like genes by Southern blot analysis.** *Eco*RI-digested BAC and cosmid DNA was hybridized with a ^{32}P-end labeled degenerate oligonucleotide to the highly conserved region spanning the catalytic histidine of serine proteases. The blot was washed in 0.1 × SSC, 0.1% SDS at 55° then exposed to film for 13 days.

David Nicol, Princess Alexandra Hospital, Brisbane, Australia. The prostate tissue samples were obtained during surgery by Dr. Robert Gardiner, Royal Brisbane Hospital, Brisbane. The ovarian cancer cell line, PEO1, was obtained from Dr. Michael McGuckin, Mater Medical Research Institute, Brisbane. Salivary gland tissue total RNA was purchased from CLONTECH and the human salivary gland cell line was obtained from Dr. Ray MacDonald, University of Texas, South-Western Medical Center, Dallas, TX. All cell lines were cultured in Dulbecco's modified Eagle's medium with 10% deactivated fetal calf serum and supplemented with 50 units/ml penicillin G and 50 μg/ml streptomycin (CSL Biosciences) and incubated at 37 °C with 5% CO_2.

RT-PCR and Southern Blot Analysis—Total RNA was extracted from the cell lines using Trizol reagent (Life Technologies, Inc.) following the manufacturer's instructions. Two μg of total RNA was reverse-transcribed into first-strand cDNA using Superscript II (Life Technologies, Inc.) in a 20-μl reaction. The cDNA was subsequently diluted 5-fold and PCR was then performed in a reaction mixture containing 1 μl of cDNA, 25 ng of each primer (Table II), 0.1 mM dNTPs (containing dATP, dCTP, dGTP, and TTP), 2 μl of 10 × reaction buffer (Roche Molecular Biochemicals), 0.5 units of platinum *Taq* (Life Technologies), and sterile water to 20 μl. The cycling conditions were 94 °C for 5 min to activate the polymerase, followed by 35 cycles of 94 °C for 1 min, 62–68 °C for 1 min (depending on the target gene), 72 °C for 1 min, and a final extension at 72 °C for 7 min. Ten μl of the PCR mixture was electrophoresed on a 2% agarose gel and visualized by ethidium bromide staining. The resulting amplicons were analyzed by Southern blot hybridization using digoxigenin 3′ end-labeled oligonucleotide probes (Table I), internal

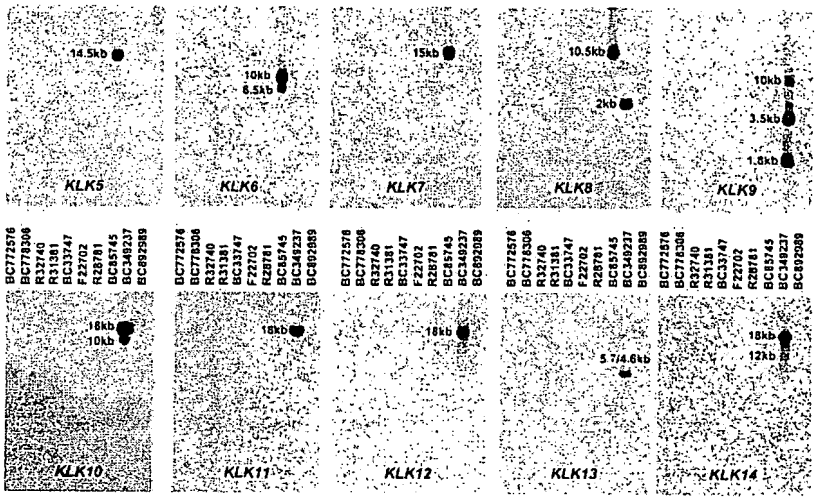

FIG. 2. **Fine mapping of *KLK5–14* to BAC and cosmid clones spanning the *KLK* locus at 19q13.4.** *Eco*RI-digested BAC and cosmid DNA was hybridized with ^{32}P-end labeled oligonucleotides or ^{32}P-random labeled cDNAs as described in the text and detailed in Table I.

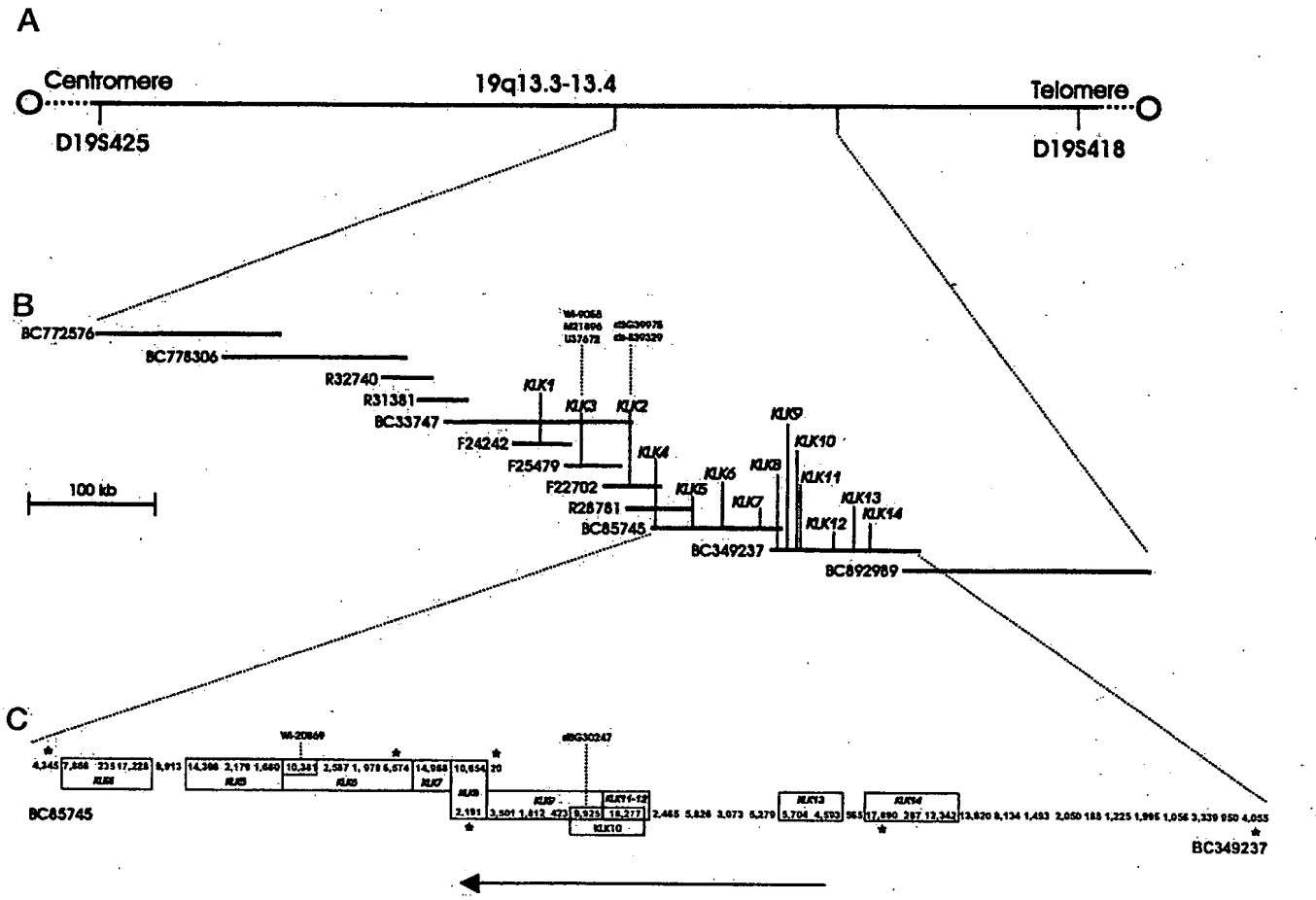

FIG. 3. **Location of the *KLK* locus at chromosome 19q13.3–13.4.** *A*, schematic representation of the interval between D19S425 and D19S418 including 19q13.3–13.4. The *KLK* locus is located proximal to D19S418. *B*, BAC and cosmid clones spanning the *KLK* locus. The position of the 14 kallikrein encoding genes are marked. *C, in silico* generated *Eco*RI digest data for BC85475 and BC349237 containing *KLK4* to *KLK14*. Fragments marked with an *asterisk* are either incomplete end fragments or have not been clearly determined because of the provisional status of the sequence. The size of fragments obtained by Southern blot analysis data (see text and Fig. 2) is consistent with the *in silico* data. The position and orientation of *KLK1* to *KLK4* on cosmids F24242, F25479, F22702, and R28781 have been reported previously (21). *KLK4* to *KLK14* are transcribed telomere to centromere as indicated by the *arrow* at the *bottom* of the figure. The position of microsatellite markers are indicated in *B* and *C*.

to the PCR primers used, in EasyHyb™ solution (Roche Molecular Biochemicals) overnight at 37 °C. Washes were performed in 0.2 × SSC, 0.1% SDS at 37 °C. In addition, for each kallikrein cDNA primer pair at least one band of the predicted size was excised from a second 2% agarose gel following electrophoresis. The DNA recovered from each band was sequenced to determine the specificity of the PCR products.

Gene-specific Oligonucleotides—Gene-specific olionucleotides, used as probes in the genomic and/or RT-PCR Southern blot analyses (Table I) were selected from regions with maximum dyshomology among the kallikreins based on a multiple sequence alignment of the 14 cDNAs. The primers used in the PCR reactions (Table II) were designed to span at least one intron for each of the 14 genes. The primer pairs were chosen for their compatibility (*i.e.* no hairpin stem loops and similar melting temperatures) and were found to be specific for the gene in question upon BLAST analysis. In addition, the percent difference for each primer/probe as compared with those of all the 14 genes was calculated and found to range from 50 to 95%. The average percent divergence for the oligonucleotides used as probes in the Southern blot experiments was 70% whereas the average for the primers used in the PCR was 74%.

TABLE III
Percentage homology between each member of the human kallikrein family at the protein level

	hK1	hK2	hK3	hK4	hK5	hK6	hK7	hK8	hK9	hK10	hK11	hK12	hK13	hK14
hK1	100	67	62	34	31	37	37	36	35	31	36	27	37	39
hK2		100	77	34	33	41	38	38	36	33	36	27	41	40
hK3			100	34	33	36	37	36	34	31	34	26	40	38
hK4				100	45	36	43	35	31	29	35	25	35	40
hK5					100	36	38	38	38	33	39	25	38	39
hK6						100	38	42	36	34	38	32	46	45
hK7							100	40	33	36	36	28	38	41
hK8								100	42	40	42	33	41	44
hK9									100	33	49	29	37	40
hK10										100	35	31	34	35
hK11											100	30	41	40
hK12												100	31	30
hK13													100	43
hK14														100

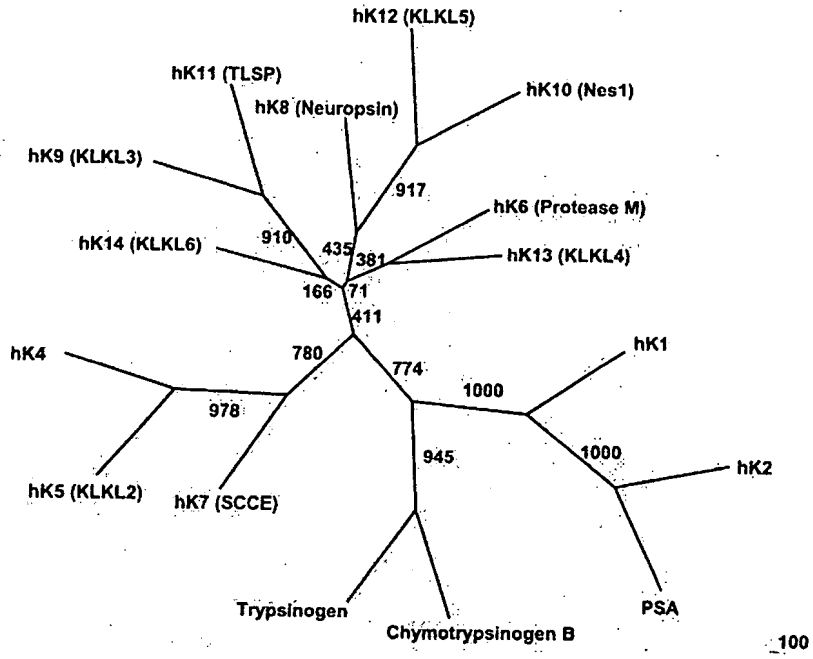

FIG. 4. **Phylogenetic tree of the expanded KLK family.** The tree was generated from an Eclustal multiple sequence alignment of the 14 kallikrein proteins using the Dayhoff program of the Phylip package.

RESULTS

Characterization of an Extended KLK Locus at Chromosome 19q13.4—To determine if there were additional kallikrein genes surrounding the original *KLK* locus on chromosome 19q13.3-13.4, Southern blot analysis was performed using degenerate oligonucleotides to consensus sequences spanning the catalytic histidine and serine residues characteristic of serine proteases. A combination of 10 BACs and cosmids, which span 400–500 kb on either side of the original *KLK* 60-kb locus, were digested with *Eco*RI and subsequently hybridized separately with each of the ^{32}P-labeled degenerate oligonucleotides. The autoradiograph resulting from hybridization with the degenerate "histidine" probe demonstrated positive signals within five of the BACs and cosmids (Fig. 1). Three of these clones (BC33747, F22702, and R28781) contain the genes, *KLK1–4* (21). However, a further two BACs (BC85745 and BC349237) also showed strong hybridization to these probes. These were not known to contain any serine protease genes. BC892989, which lies telomeric to the *KLK* cluster and the 4 clones centromeric to the cluster (R31381, R32740, BC778306, and BC772576) did not produce any positive signals even though the membrane was exposed for 13 days. The degenerate "serine" oligonucleotide probe could not be used successfully in these experiments, probably because of its degeneracy (512-fold). Thus, additional serine protease genes are only found within a region encompassed by BACs BC85745 and BC349237, and there are no other serine protease genes approximately 200 kb telomeric of BC349237 and 400 kb centromeric of the original *KLK*1–3 cluster.

In order to identify the serine protease genes which were detected using the degenerate probe described above, the same set of *Eco*RI-digested BACs and cosmids were examined using one or two gene-specific oligonucleotides or a cDNA probe to each of the *KLK*5-*14* genes. The *KLK5* probe hybridized to a 14.5-kb fragment on BC85745 (Fig. 2). Positive signals were detected on BC85745 at 10 and 6.5 kb for *KLK6* and at 15 kb for *KLK7*. A signal was observed for *KLK8* on both BC85745 and BC349237 at 10.5 and 2 kb, respectively, and *KLK9* was found to be located within three of the *Eco*RI fragments at 10, 3.5, and 1.8 kb on BC349237. Two bands at 18 and 10 kb were detected for *KLK10* on BC349237 whereas only the 18-kb band was positive for *KLK11*. The remaining three *KLK*-like genes were all present on BC349237. *KLK12* was present on an 18-kb band. Although only one signal (at approximately 5 kb) was detected for *KLK13*, this band represents two individual *Eco*RI fragments of 5.7 and 4.6 kb in size. The *KLK14* cDNA probe hybridized to an 18- and a 12-kb band.

Concurrent with the physical mapping studies, computer analysis was performed on approximately 320 kb of draft genomic sequence spanning the *KLK* locus on chromosome 19q.

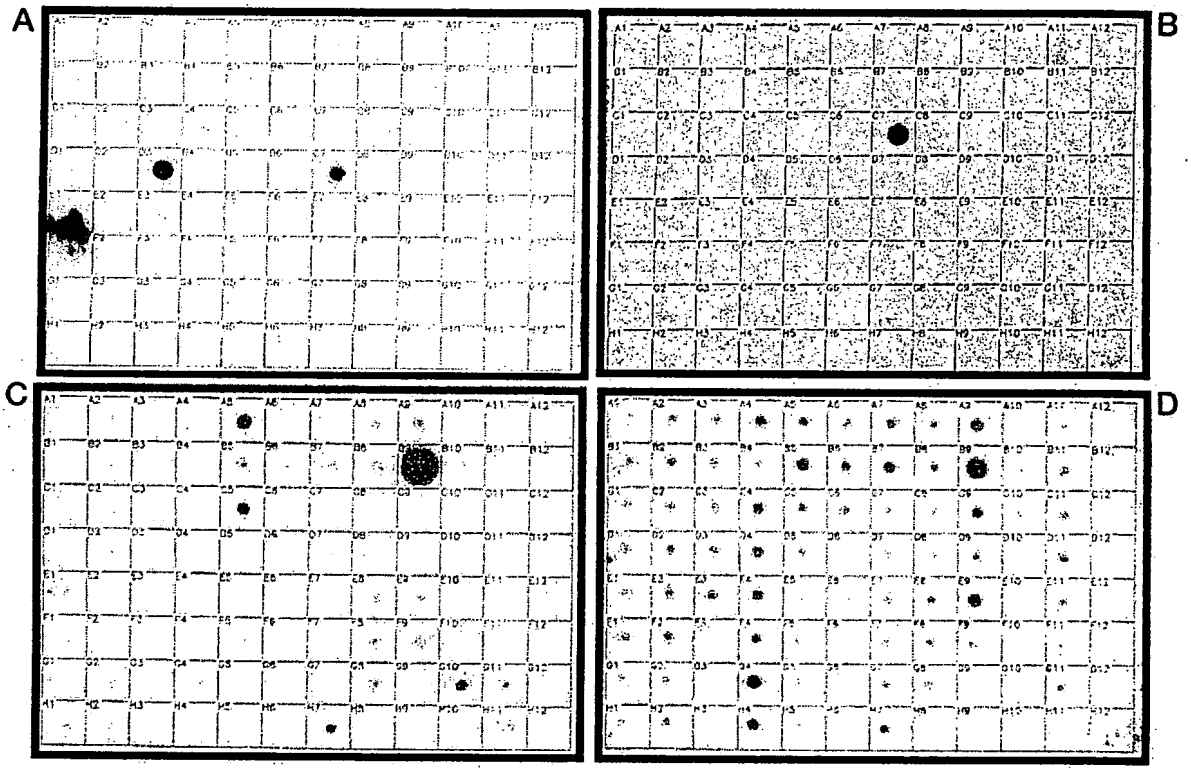

FIG. 5. **mRNA expression patterns.** Membranes containing poly(A)$^+$ from human tissues and cell lines were screened with ^{32}P-labeled probes to each of the 14 *KLK* genes. Autoradiographs representative of the patterns obtained are shown. *A, KLK1. B, KLK2–4. C, KLK10. D, KLK11.* Tissues arrayed on the membranes in *panels A* and *B*: A1, whole brain; A2, amygdala; A3, caudate nucleus; A4, cerebellum; A5, cerebral cortex; A6, frontal lobe; A7, hippocampus; A8, medulla oblongata; B1, occipital lobe; B2, putamen; B3, substantia nigra; B4, temporal lobe; B5, thalamus; B6, nucleus accumbeus; B7, spinal cord; C1, heart; C2, aorta; C3, skeletal muscle; C4, skeletal muscle; C5, bladder; C6, uterus; C7, prostate; C8, stomach; D1, testis; D2, ovary; D3, pancreas; D4 pituitary gland; D5 adrenal gland; D6, thyroid gland; D7, salivary gland; D8, mammary gland; E1, kidney; E2, liver; E3, small intestine; E4, spleen; E5, thymus; E6, peripheral blood leukocyte; E7, lymph node; E8, bone marrow; F1, appendix; F2, lung; F3, trachea; F4, placenta; G1, fetal brain; G2, fetal heart; G3, fetal kidney; G4, fetal liver; G5, fetal spleen; G6, fetal thymus; G7, fetal lung; H1, yeast total RNA, 100 ng; H2, yeast tRNA, 100 ng; H3, *Escherichia coli* rRNA, 100 ng; H4, *E. coli* DNA, 100 ng; H5, Poly r(A) 100 ng; H6, human C_0t-1 DNA, 100 ng; H7, human DNA, 100 ng; H8, human DNA, 500 ng. Tissues arrayed on the membranes in *panels C* and *D*: A1, whole brain; A2, cerebellum, left; A3, substantia nigra; A4, heart; A5, esophagus; A6, colon; A7, kidney; A8, lung; A9, liver; A10, leukemia, HL-60; A11, fetal brain; A12, yeast total RNA; B1, cerebral cortex; B2, cerebellum, right; B3, accumbens nucleus; B4, aorta; B5, stomach; B6, colon, descending; B7, skeletal muscle; B8, placenta; B9, pancreas; B10, HeLa S3; B11, fetal heart; B12, yeast tRNA; C1, frontal lobe; C2, corpus callosum; C3, thalamus; C4, atrium, left; C5, duodenum; C6, rectum; C7, spleen; C8, bladder; C9, adrenal gland; C10, leukemia, K-562; C11, fetal kidney; C12, *E. coli* rRNA; D1, parietal lobe; D2, amygdala; D3, pituitary gland; D4 atrium, right; D5 jejunum; D7, thymus; D8, uterus; D9, thyroid gland; D10, leukemia MOLT-4; D11, fetal liver; D12, *E. coli* DNA; E1, occipital lobe; E2, caudate nucleus; E3, spinal cord; E4, ventrical, left; E5, ileum; E7, peripheral blood leukocyte; E8, prostate; E9, salivary gland; E10, Burkitt's lymphoma, Raji; E11, fetal spleen; E12, Poly r(A); F1, temporal lobe; F2, hippocampus; F4, ventrical, right; F5, ilocecum; F7, lymph node; F8, testis; F9, mammary gland; F10, Burkitt's lymphoma, Daudi; F11, fetal thymus; F12, human C_0t-1 DNA; G1, cerebral cortex; G2, medulla oblongata; G4 interventricular septum; G5, appendix; G7, bone marrow; G8, ovary; G10, colorectal adenocarcinoma, SW480; G11, fetal lung; G12, human DNA 100 ng; H1, pons; H2, putamen; H4, apex of heart; H5, colon, ascending; H7, trachea; H10, lung carcinoma, A549; H12, human DNA 500 ng.

To identify novel protease genes in this region, searches of the genomic sequence were performed using the consensus sequences spanning each of the histidine, aspartate, and serine residues necessary for the enzymatic activity of serine proteases. These searches identified 10 potential serine protease genes in addition to known kallikrein genes, *KLK1–4*. Further searches were then performed with the sequence for each of the *KLK*-like serine protease genes: Protease M (*KLK6*), SCCE (*KLK7*), neuropsin (*KLK8*), Nes1 (*KLK10*), TLSP (*KLK11*), KLKL2 (*KLK5*), KLKL3 (*KLK9*), KLKL4 (*KLK13*), KLKL5 (*KLK12*), and KLKL6 (*KLK14*), using the BLASTN algorithm. All 10 of these genes were identified within this sequence and were found to correspond to the genes identified from searches using the three serine protease consensus sequences. The genes were all located in a telomeric position with respect to the *KLK2–4* genes. The *KLK1* gene is located ~31 kb centromeric of *KLK2–4* (21, 22); sequence data is not available for this region. Also, an *Eco*RI restriction map of the sequence contained within BACs BC85745 and BC349237 was constructed *in silico*. This map was used in conjunction with the physical mapping data and the *Eco*RI map available at the Lawrence Livermore National Laboratory website, to determine the exact position of each gene and to reconstruct a definitive *Eco*RI map of this region (Fig. 3). The size of each of the bands identified by Southern blot analysis (Fig. 2) agreed with the *in silico* generated *Eco*RI fragment sizes. From the map, the order of the expanded *KLK* locus (according to the old nomenclature) is centromere-*KLK1-KLK3-KLK2-KLK4*-KLKL2-Protease M-SCCE-neuropsin-KLKL3-Nes1-TLSP-KLKL5-KLKL4-KLKL6-telomere (new nomenclature: centromere-*KLK1-KLK3-KLK2-KLK4*- to -*KLK14*-telomere). Analysis of the contiguous sequence indicated that *KLK5* to *KLK14* are transcribed telomere to centromere (Fig. 3).

Clarification of the Position of Microsatellite Markers within the KLK Locus at Proximal 19q13.4—To clarify the order and position of the microsatellite markers within the *KLK* locus, microsatellite marker sequences were aligned against the chromosome 19q proximal 13.4 draft sequence data using the BLASTN algorithm. For this purpose, a data base of contiguous sequence was established using the resources available at the

TABLE IV
Summary of mRNA expression patterns
Membranes containing poly(A)$^+$ from human tissues and cell lines were screened with ^{32}P random labeled probes to each of the 14 kallikreins.

Gene	High expression	Moderate expression	Gene	High expression	Moderate expression
KLK1	Kidney Pancreas Salivary gland		*KLK8* (Neuropsin)	Pancreas	Duodenum Liver Esophagus Stomach Salivary gland
KLK2	Prostate		*KLK9* (KLKL3)	Pancreas	
KLK3	Prostate		*KLK10* (Nes1)	Pancreas	Esophagus Duodenum Trachea Colorectal adenocarcinoma
KLK4	Prostate		*KLK11* (TLSP)	Pancreas	Heart Salivary gland Stomach Liver Skeletal muscle
KLK5 (KLKL2)	Testis Mammary gland	Salivary gland Esophagus	*KLK12* (KLKL5)	Pancreas	Brain Duodenum Appendix
KLK6 (Protease M)	Pancreas	Spinal cord Brain Testis Appendix Colorectal adenocarcinoma	*KLK13* (KLKL4)	Pancreas Esophagus Appendix	Duodenum Stomach Brain Testis Prostate Salivary gland
KLK7 (SCCE)	Pancreas		*KLK14* (KLKL6)	Brain Bone marrow Fetal liver	Liver Pancreas Fetal spleen Prostate

Australian National Genome Information Service website. Marker sequences between 272.77 and 276.46cR$_{3000}$ at interval D19S425-D19S418 on chromosome 19 were aligned with the nucleotide sequence within the data base. The data obtained from these searches was used to re-evaluate the order and presence of these markers within the region of the *KLK* locus (Fig. 3). Markers WI-9055 and M21896, previously recognized as partial *KLK3* sequence, as well as U37672 (previously unknown) were all found to lie within the *KLK3* sequence on cosmid F25479. The two markers previously identified as partial *KLK1* sequence, stSG39975 and sts-S39329, are actually partial sequences of the *KLK2* gene and were found within the sequence of R33359 which overlaps cosmid F22702 at this location. The marker, WI-20869, recognized as *KLK6* partial sequence, was confirmed to lie within the *KLK6* coding sequence on BC85745. Similarly, stSG30247 was shown to lie within the *KLK10* sequence on BC349237. In contrast, none of the remaining microsatellite markers listed as being located in the interval D19S425-D19S418 were detected within the available genomic sequence of the extended *KLK* locus. These markers are stSG32045 (EST), stSG1457 (β-Electron transfer flavoprotein), stSG3375 (OTK18 mRNA), stSG372 (CD33 antigen), sts-H71236 (EST), SGC35527 (granulocyte colony stimulating factor-induced gene), U37672 (unknown), stSG30429 (EST), sts-AA028917 (EST, highly similar to zinc finger protein 91), L12214 (unknown). Prior mapping work shows genes *CD33* and *ETFB* to be distal to the KLK locus by ~40 and ~250 kb, respectively.

Divergence of Serine Protease Genes within the Extended KLK Locus—Protein alignments were performed to establish the degree of similarity and divergence of the 14 serine protease genes located within the expanded *KLK* locus. The percentage homology between the proteins encoded by the 14 *KLK* genes was also calculated (Table III). The classical *KLK* family members, hK1,[3] hK2 and PSA (encoded by *KLK1, KLK2,* and *KLK3,* respectively), display a high degree of homology (62–77% identity) at the protein level. hK4, is less similar with only 34% identity to PSA. As shown in Table III, hK4 possesses a similar degree of homology to the more recently identified serine proteases hK5-14 (25–45%) as it does to PSA. Similarly, the degree of homology between the protease groups, hK5–14 and hK1–3, ranges from 26 to 41%. Phylogenetic analysis of the 14 protein sequences produced similar results to the homology studies (Fig. 4). The phylogenetic tree demonstrated that hK1, hK2, and PSA lie on one branch, as do hK4, hK5, and hK7. The two serine proteases, hK10 and hK12, were clustered on a third branch while hK9 and hK11 were found on a fourth branch. However, the remaining four members (hK6, hK8, hK13, and hK14) were located around the base of the tree. Due to the low branch numbers assigned to these four kallikreins, it is difficult to precisely determine their degree of divergence from the family as a whole. Nevertheless, the high branch numbers observed overall suggest that this expanded family of human kallikreins consist of at least four subfamilies.

mRNA Expression Patterns—To establish a comprehensive pattern of expression for the 11 *KLK* genes, *KLK4–14,* as compared with that known for *KLK1–3,* arrays containing poly(A)$^+$ RNA from either 50 or 76 different human tissues and cell lines were screened using ^{32}P-labeled probes to each of the 14 *KLK* serine protease genes. Examples of the autoradiographs obtained are shown in Fig. 5 and a summary of the expression levels is shown in Table IV. The expression patterns observed for *KLK1–4* were as previously reported, where *KLK1* was detected in the kidney, pancreas, and salivary gland while *KLK2–4* were present only within the prostate (27, 32, 33). High expression of the *KLK10* gene was observed in the pancreas and moderate expression in the esophagus, duodenum, trachea, and colorectal adenocarcinoma. Low expression was seen in tissues such as the liver, kidney, ovary, testis, prostate, and salivary gland. High expression of *KLK11* was also de-

[3] Kallikrein family protein nomenclature used here is that described by Berg *et al.* (3), *e.g.* protein hK1 corresponds to gene *KLK1.*

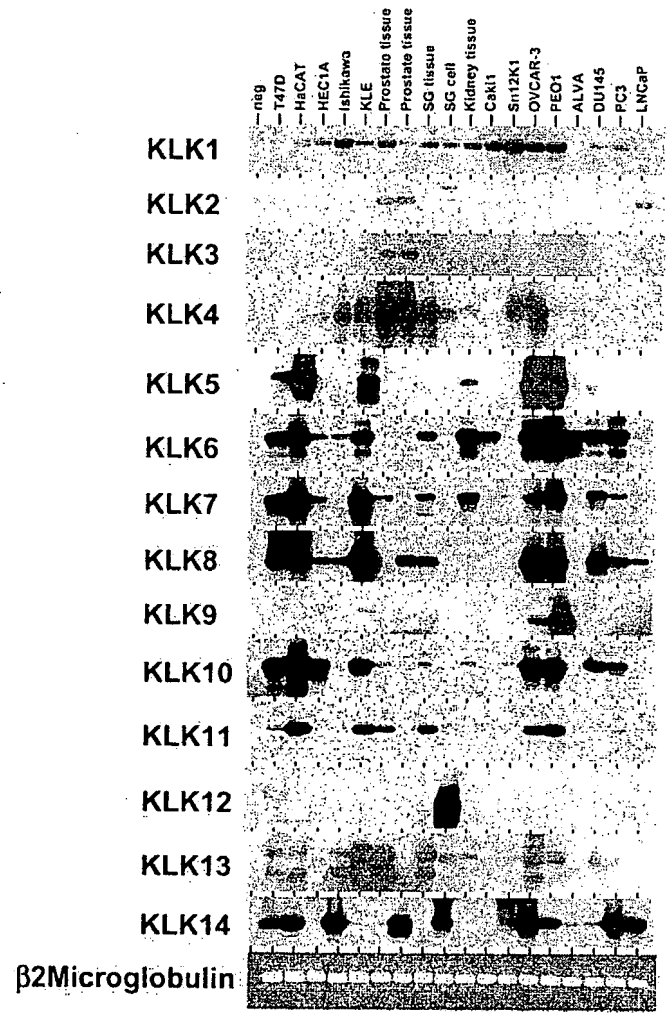

FIG. 6. **Expression pattern analysis by RT-PCR.** Analysis was performed on total RNA isolated from the displayed cell lines and tissues using the primers listed in Table II. RT-PCR products were transferred to nylon membranes and probed by Southern blot analysis using digoxigenin-labeled oligonucleotide probes (Table I) internal to the PCR primers used. SG, salivary gland.

tected in the pancreas. Moderate expression of *KLK11* was observed in the heart, salivary gland, stomach, liver, and skeletal muscle, while low expression was observed in most of the remaining tissues represented on the blot. Of interest, is the overall high expression of nearly all of these new *KLK* genes in the pancreas, particularly *KLK6* to *KLK13*. Furthermore, moderate expression was observed for several genes in the salivary gland, esophagus, trachea, brain, gastrointestinal, and reproductive organs.

In addition, RT-PCR was performed using total RNA extracted from a number of different transformed and cancer cell lines as well as normal and cancer tissues. To increase the sensitivity of detection, Southern blot analysis using gene-specific oligonucleotide probes was performed on the PCR products as shown in Fig. 6 and summarized in Table V. To confirm the specificity of the PCR for each of the 14 genes, bands of the predicted size were sequenced and found to correspond to the gene of interest (see Table V). From these results, it can be seen that each gene displays a distinctive expression pattern with several genes (*KLK1, KLK6, KLK7, KLK9, KLK10, KLK11, KLK13,* and *KLK14*) being expressed in a broad range of tissues/cell lines whereas other genes (*KLK2–5, KLK8,* and *KLK12*) are less widely expressed. Of interest is the high expression of many of the newer *KLK* genes (*KLK5–14*) in keratinocyte (HaCAT), ovarian (OVCAR-3 and PEO1), breast (T47D), and endometrial (KLE) cancer cell lines. In addition, all genes (except for *KLK12*) appear to have varying levels of expression in one of the four prostate cancer cell lines. Furthermore, several of these genes are moderately expressed in the salivary gland and kidney. Additional bands were observed for several of the genes being examined. Some of these alternate bands were sequenced and found to represent variants of the gene in question (data not shown). Similarly, other studies have identified alternatively spliced transcripts for these new *KLK* genes (19).

DISCUSSION

The tissue kallikrein genes, particularly those in rodents, represent a distinctive family of a large number of genes, which are highly conserved in sequence and genomic organization. Nevertheless, the encoded enzymes exhibit specific tissue expression patterns and diverse biological functions. We have now determined that the size of the human *KLK* gene family, like its rodent counterparts, is large and it is likely to consist of 14 genes. Physical mapping of these genes at chromosome 19q13.4 concurs with recent computer analysis of draft sequence data of this region (30). The tissue-specific patterns of expression observed for these 14 genes shows, again like rodents, a diverse expression profile for individual members of the family. Of interest, however, is that the human family has diverged further in sequence.

The rodent *KLK* families are very large with 13 genes in the rat and 24 genes in the mouse (23–25) and both cluster to one chromosomal loci (23, 25, 26). The original human *KLK* family (*KLK1–3*), also clustered to one loci (21, 22), has a high degree of similarity to the rodent family with 65–72% identity at the nucleotide level. Due to the similarity between these *KLK* families and the previous report by Murray *et al.* (29), which proposed the existence of up to 19 human *KLK* genes by screening human genomic DNA with a monkey cDNA, we began the search for new members of the human *KLK* family. In addition, previous studies placed the genes encoding the serine proteases Protease M and Nes1 (*KLK6* and *KLK10* respectively), on chromosome 19q in a region to which the original kallikreins, *KLK1–3*, also map (10, 13).

Our initial studies involving Southern blot analysis identified several potentially novel serine protease genes within the surrounding regions of chromosome 19q. Further investigation of these results identified five published serine protease genes (previously named Protease M, Nes1, TLSP, SCCE, and Neuropsin) on the BACs and cosmids containing the human *KLK* genes, *KLK2-4*. All five genes were detected telomeric to *KLK4* and thus were considered to be candidates for new members of the human *KLK* family. Furthermore, several other *KLK*-related sequences, *KLK5, KLK9,* and *KLK12–14*, were also physically mapped to 19q13.4. This confirmed the recently published data identifying the position of these genes on chromosome 19, by computer analysis (30). In addition, as the data obtained by Southern blot analysis with a degenerate probe to the histidine-encoding region of serine proteases, detected unaccounted for signals only within the two BACs adjacent to the classical *KLK* locus (BC85745 and BC349237) and no signal from DNA, both telomeric to these two BACs or centromeric to the original *KLK* cluster, it would appear that there are only 10 additional serine protease genes within the surrounding regions of the classical *KLK* locus. In accordance with our data, the next gene telomeric from *KLK14* (KLK-L6) is an unrelated gene, OB-BP-like, and encodes a putative leptin-binding protein (34). Accordingly, although sequence is not yet available for the region centromeric to *KLK1*, it is likely the complete *KLK* locus has now been identified.

TABLE V

Expression patterns of the 14 KLK genes obtained by RT-PCR analysis

The following symbols used represent: +++, high expression; ++, moderate expression; +, low expression; (◆) PCR products of the predicted size sequenced and confirmed to be the correct sequence.

Cell line	*KLK1*	*KLK2*	*KLK3*	*KLK4*	*KLK5* KLKL2	*KLK6* protease M	*KLK7* SCCE	*KLK8* neuropsin	*KLK9* KLKL3	*KLK10* Nes1	*KLK11* TLSP	*KLK12* KLKL5	*KLK13* KLKL4	*KLK14* KLKL6
T47D					++	+++	+++	++◆	+	+++	++	+	++	+++
HaCAT	+			+	+++	+++	+++◆	++		+++◆	+++◆		++	+++
HEC1A	+++◆			+		++	++	+		+++◆	+			
Ishikawa	+++			++◆		++		+					++	+++
KLE	++			++	+++	+++	+++	+++	+	+++	+++		+++	+
Prostate tissue	++	+++◆	+++◆	+++◆			++			+	+++		+++	+
Prostate tissue	+	+++	+++	+++				++			+			+++
SG tissue	++		+	++		++	++	++		+	+++		++	
SG cell	++	+		++		+						+++◆	++	+++
Kidney tissue	+++◆			+	++	+++◆	++			+	+		+	
Caki1	+++					+++								
Sn12K1	+++			++			+			+			+	+
OVCAR-3	+++			+++	+++◆	+++	++	+++◆	++◆	+++	+++		++◆	+++◆
PEO1	+++			+	+++	+++◆	+++◆	+++	+++	+++◆	+++◆		++	+++
ALVA				+		+++				+				+
DU145	+				+	+++	+++	++		++			+	+
PC3	+				+	+++	++	++		++	+			+++
LNCaP	+	++	+					+						+++

Comparison of the rat *KLK* gene sequences indicates concerted evolution of the family that results in sequence homogenization of all family members (35). In addition, there is a clear conservation of genomic organization with 5 coding exons for all *KLK* genes, although some of the newer family members have additional exons containing 5′-untranslated sequences (6, 7, 10, 13, 15–21, 23, 27). Characterization of the mouse and rat *KLK* genes showed that nucleotide sequence identity exceeds 85% and extends throughout the coding regions, introns, and gene flanking sequences (23, 24, 35). The first characterized human *KLK* family members, *KLK1–3*, also show high sequence homology, at both the nucleotide and protein level (62–77%) (6, 7, 23, 32). However, in man the level of homology drops sharply for kallikrein genes telomeric of *KLK1–3*. The next closest gene along the locus, *KLK4* (15, 21), is 34% identical to *KLK3* (15, 21, 27). Similarly, as reported here and by others, the newer members of the family (*KLK5–14*) are less conserved with 25–49% similarity at the protein level, both between each other and with respect to *KLK1-KLK4* (10, 13, 15–20). This divergence was also evident from our phylogenetic analysis. The degree of divergence seen within the human *KLK* family tree suggests that there are a number of subgroups within this species. The mouse *Klk* and original human *KLK* gene families were shown to lie in syntenic regions on chromosomes 7 and 19q, respectively (21–23). Of interest, the putative mouse orthologue of one of the newer human *KLK* genes, *KLK8* (neuropsin), also maps to this region on chromosome 7 (36).

Analysis of microsatellite markers within the *KLK* locus showed that the current order needed to be revised. Alignment of each of the markers with the genomic sequence for chromosome 19q revealed that a number of the genes listed were in fact not found within the sequence available spanning the *KLK* locus. Nevertheless, those markers representative of partial *KLK2*, *KLK3*, *KLK6*, and *KLK10* gene sequences were indeed detected within the genomic sequence. Thus, the listing of the microsatellite markers within the *KLK* locus has now been further clarified.

Poly(A)$^+$ RNA dot blot analysis indicated the prostate-restricted expression of *KLK2–4* and abundant expression of *KLK6–13* in pancreas. Of interest, in an analogous fashion to the mouse and rat clustered *Klk*/*KLK* genes which are predominantly expressed in the salivary gland (9, 23), *KLK2–4* are clustered together within the human locus as are *KLK6–13*. Furthermore, in man the two genes displaying high expression in the testis/mammary gland and brain/bone marrow/fetal liver, *KLK5* and *KLK14*, respectively, are located either side of the cluster of pancreatic-expressed genes. As the *KLK5* gene lies adjacent to the *KLK2–4* cluster, also found in a hormone-dependent tissue, these four genes may also share common transcriptional control elements. As has been suggested for the rat kallikreins (9, 24, 35), this clustering of genes having common restricted expression patterns likely indicates evolutionary conservation of elements conferring tissue specificity such as duplicated promoter elements or a locus control region. Whether *KLK* transcriptional control in man is conferred through either or both of these elements is not yet known. Certainly our phylogenetic studies indicate that the human kallikreins have diverged to a much greater extent than the comparably sized rodent families. This observation may indicate that duplicated promoter elements would not be expected to have survived the apparent evolutionary pressure to diverge in man or, alternatively, may account for variations in expression patterns. Of interest, a pancreatic-specific-like sequence, similar to that identified in other rat pancreatic serine proteases (37, 38), has been identified in the promoter of the rat, mouse, and human *KLK1* genes (24). In contrast, duplicated promoter regions have already been identified in the *KLK2* and *KLK3* genes (39, 40). In addition, although expression profiles of rat kallikreins show that individual genes are expressed in diverse tissue-specific patterns there is still high conservation of potential transcriptional regulatory regions (24, 25).

Our RT-PCR analysis of the cell line expression patterns of *KLK5–14* indicated that most of these genes display significant expression in keratinocytes, breast, ovarian, and endometrial cancer cell lines and in salivary gland tissue and a salivary gland cell line. Expression of these *KLK*s in cancer-derived and transformed cell lines is not surprising, and concurs with other studies (10, 11, 13–21, 27, 34, 43), as proteases are well known to aid in malignant processes such as invasion, metastasis, and proliferation (41, 42). It will be of interest to determine whether any of the kallikreins are up-regulated in tumors and the role of these proteolytic enzymes in tumor progression.

Although the enzymic function of these newer kallikrein enzymes is yet to be elucidated, an indication of the tissue in which they will be functional may be gained from the patterns of expression noted above. From several *in vitro* biochemical studies, PSA and hK2 are known to degrade extracellular matrix proteins and PSA degrades the insulin-like growth factor-binding protein 3 to regulate the bio-availability of insulin-like growth factor 1; a key factor in cancer cell proliferation (8, 44).

The hK2 enzyme is likely the physiological activator of PSA (8) and hK2 can also activate urokinase plasminogen activator (45), another key enzyme in matrix degradation, demonstrating that involvement in an enzymic cascade may also be an important function of this family of enzymes. Both PSA and hK2 are useful diagnostic and prognostic markers in prostate and breast cancer (8, 34). Thus, many of the newer members of the *KLK* family which are all expressed in a variety of cancer cell lines may have similar functional roles in these diseases. Similarly, other new *KLK* members, which are highly expressed in skin and brain, have been suggested to play important roles in the (patho)physiology of these organs (12, 46–48). The high pancreatic and gastrointestinal tract expression of several *KLK* genes may also prove to be of significance since tissue kallikrein has already been shown to play a role in the (patho)physiology of these tissues (2, 49).

In summary, we have physically mapped the *KLK* locus, providing a definitive fine map of this region at 19q13.4 and have further clarified the size and expression patterns of the extended human *KLK* gene family. Our tissue mRNA expression studies indicate *KLK* expression can be largely delineated between those having prostate restricted expression (*KLK2–4*) and those having abundant pancreas expression (*KLK6–13*). These profiles suggest evolutionary conservation of elements conferring tissue specificity such as duplicated promoter elements and locus control regions. Future studies will be directed at characterizing these elements and in elucidating the roles of kallikrein serine proteases in human biology and disease.

Acknowledgments—We thank Drs. David Nicol (Princess Alexandra Hospital, Brisbane), Rosemary Hall (Flinders Cancer Center, Adelaide), Mark Hayes (Endocrine and Diabetes Department, Royal Children's Hospital, Brisbane), Michael McGuckin (Mater Medical Research Institute, Brisbane), A. Nakla (St Luke's Hospital, New York), Robert Gardiner (Royal Brisbane Hospital, Brisbane), and Ray MacDonald (University of Texas, Southwestern Medical Center, Dallas, TX) for their generous gift of the different cell lines or tissues used in this study. Furthermore, we thank Dr. Ying Dong, Fiona Rae, Tara Veveris-Lowe, and Rachael Collard for propagating cell lines and providing RNA. We also thank Fiona Rae and Aneel Kaushal for assistance with some of the hybridizations.

REFERENCES

1. Clements, J. (1997) in *The Kinin System* (Farmer, S., ed) pp. 71–97, Academic Press, Inc., New York
2. Bhoola, K. D., Figueroa, C. D., and Worthy, K. (1992) *Pharmacol. Rev.* **44**, 1–80
3. Berg, T., Bradshaw, R. A., Carretero, O. A., Chao, J., Chao, L., Clements, J. A., Fahnestock, M., Fritz, H., Gauthier, F., MacDonald, R. J., Margolius, H. S., Morris, B. J., Richards, R. I., and Scicli, A. G. (1992) *Agents Actions* **38**, 19–25
4. Drinkwater, C. C., Evans, B. A., and Richards, R. I. (1988) *Trends Biochem. Sci.* **13**, 169–72
5. Clements, J. A. (1989) *Endocr. Rev.* **10**, 393–419
6. Riegman, P. H. J., Vlietstra, R. J., van der Korput, J. A. G. M., Romijin, J. C., and Trapman, J. (1989) *Biochem. Cell Biol.* **159**, 95–102
7. Schedlich, L. J., Bennetts, B. H., and Morris, B. J. (1987) *DNA* **6**, 429–437
8. Rittenhouse, H. G., Finlay, J. A., Mikolajczyk, S. D., and Partin, A. W. (1998) *Crit. Rev. Clin. Lab. Sci.* **35**, 275–368
9. MacDonald, R. J., Southard-Smith, E. M., and Kroon, E. (1996) *J. Biol. Chem.* **271**, 13684–13690
10. Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996) *Mol. Med.* **2**, 624–636
11. Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, T., Nakazato, H., and Yamaguchi, N. (1997) *Biochim. Biophys. Acta* **1350**, 11–14
12. Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., Mackellar, W., Hepburn, D., Corvalan, J., McClure, D., Liu, X., Stephenson, D., Clements, J., and Johnstone, E. (1997) *J. Biol. Chem.* **272**, 25135–25142
13. Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996) *Cancer Res.* **56**, 3371–3379
14. Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998) *Cancer Res.* **58**, 4782–4786
15. Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. (1999) *Cancer Res.* **59**, 4252–4256
16. Yousef, G. M., and Diamandis, E. P. (1999) *J. Biol. Chem.* **274**, 37511–37516
17. Yousef, G. M., Luo, L. Y., Scherer, S. W., Sotiropoulou, G., and Diamandis, E. P. (1999) *Genomics* **62**, 251–259
18. Yousef, G. M., Scorilas, A., and Diamandis, E. P. (2000) *Genomics* **63**, 88–96
19. Yousef, G. M., Chang, A., and Diamandis, E. P. (2000) *J. Biol. Chem.* **275**, 11891–11898
20. Yousef, G. M., and Diamandis, E. P. (2000) *Genomics* **65**, 184–194
21. Stephenson, S. A., Verity, K., Ashworth, L. K., and Clements, J. A. (1999) *J. Biol. Chem.* **274**, 23210–23214
22. Riegman, P. H., Vlietstra, R. J., Suumeijer, L., Cleutjens, C. B., and Trapman, J. (1992) *Genomics* **14**, 6–11
23. Evans, B. A., Drinkwater, C. C., and Richards, R. I. (1987) *J. Biol. Chem.* **262**, 8027–8034
24. Wines, D. R., Brady, J. M., Pritchett, D. B., Roberts, J. L., and MacDonald, R. J. (1989) *J. Biol. Chem.* **264**, 7653–62
25. Southard-Smith, M., Pierce, J. C., and MacDonald, R. J. (1994) *Genomics* **22**, 404–417
26. Kurtz, T. W., and St. Lezin, E. M. (1992) *J. Am. Soc. Nephrol.* **3**, 28–34
27. Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3114–3119
28. Lin, F. K., Lin, C. H., Chou, C. C., Chen, K., Lu, H. S., Bacheller, W., Herrera, C., Jones, T., Chao, J., and Chao, L. (1993) *Biochim. Biophys. Acta* **1173**, 325–328
29. Murray, S. R., Chao, J., Lin, F.-K., and Chao, L. (1990) *J. Cardiovasc. Pharmacol.* **15**, S7–S16
30. Yousef, G. M., Luo, L. Y., and Diamandis, E. P. (1999) *Anticancer Res.* **19**, 2843–2852
31. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22**, 521–565
32. Fukushima, D., Kitamura, N., and Nakanishi, S. (1985) *Biochemistry* **24**, 8037–8043
33. Chapdelaine, P., Paradis, G., Tremblay, R. R., and Dube, J. Y. (1988) *FEBS Lett.* **236**, 205–208
34. Diamandis, E. P., Yousef, G. M., Luo, L-Y., Magklara, A., and Obiezu, C. V. (2000) *Trends Endocrinol. Metab.* **11**, 54–60
35. Wines, D. R., Brady, J. M., Southard, E. M., and MacDonald, R. J. (1991) *J. Mol. Evol.* **32**, 476–492
36. Chen, Z. L., Yoshida, S., Kato, K., Momota, Y., Suzuki, J., Tanaka, T., Ito, J., Nishino, H., Aimoto, S., Kiyama, H., and Shiosaka, S. (1995) *J. Neurosci.* **15**, 5088–5097
37. Swift, G. H., Craik, C. S., Stary, S. J., Quinto, C., Lahaie, R. G., Rutter, W. J., and MacDonald, R. J. (1984) *J. Biol. Chem.* **259**, 14271–14278
38. Boulet, A. M., Erwin, C. R., and Rutter, W. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3599–3602
39. Wolf, D. A., Schulz, P., and Fittler, F. (1992) *Mol. Endocrinol.* **6**, 753–762
40. Riegman, P. H. J., Vlietstra, R. J., Van der Korput, J. A. G. M., Brinkmann, A. O., and Trapman, J. (1991) *Mol. Endocrinol.* **5**, 1921–1930
41. Carroll, V. A., and Binder, B. R. (1999) *Semin. Thromb. Hemostasis* **25**, 183–97
42. DeClerck, Y. A., Imren, S., Montgomery, A. M., Mueller, B. M., Reisfeld, R. A., and Laug, W. E. (1997) *Adv. Exp. Med. Biol.* **425**, 89–97
43. Underwood, L. J., Tanimoto, H., Wang, Y., Shigemasa, K., Parmley, T. H., and O'Brien, T. J. (1999) *Cancer Res.* **59**, 4435–4439
44. Cohen, P., Graves, H. C. B., Peehl, D. M., Kamerei, M., Guidice, L. C., and Rosenfeld, R. G. (1992) *J. Clin. Endocrinol. Metab.* **75**, 1046–1053
45. Frenette, G., Tremblay, R. R., Lazure, C., and Dube, J. Y. (1997) *Int. J. Cancer* **71**, 897–899
46. Lundstrom, A., and Egelrud, T. (1991) *Acta Dermato-Venereol.* **71**, 471–474
47. Yoshida, S., Taniguchi, M., Suemoto, T., Oka, T., He, X. P., and Shiosaka, S. (1998) *Biochim. Biophys. Acta* **1399**, 225–228
48. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998) *Gene (Amst.)* **213**, 9–16
49. Griebacher, T., and Lembeck, F. (1997) in *The Kinin System* (Farmer, S., ed) pp. 197–217, Academic Press, Inc., New York
50. Diamandis, E. P., Yousef, G. M., Clements, J. A., Ashworth, L., Yoshida, S., Egelrud, T., Nelson, P. S., Shiosaka, S., Band, V., Little, S., Lilja, H., Stenman, U.-H., and Wain, H. (2000) *Clin. Chem.* in press